Volume 2, Issue 10, October 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Data Mining Techniques

Kalyani M Raval (B.Com, MSc IT) *
Lecturer in B.Com MIP and PGDCA
M. J. College of Commerce.
Maharaja Krishnakumarsinhji Bhavnagar University
Bhavnagar [Gujarat – India]

Abstract— Data mining is a process which finds useful patterns from large amount of data The process of extracting previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions - Simoudis 1996 This data mining definition has business flavor and for business environments. However, data mining is a process that can be applied to any type of data ranging from weather forecasting, electric load prediction, product design, etc. Data mining also can be defined as the computer-aid process that digs and analyzes enormous sets of data and then extracting the knowledge or information out of it. By its simplest definition, data mining automates the detections of relevant patterns in database.

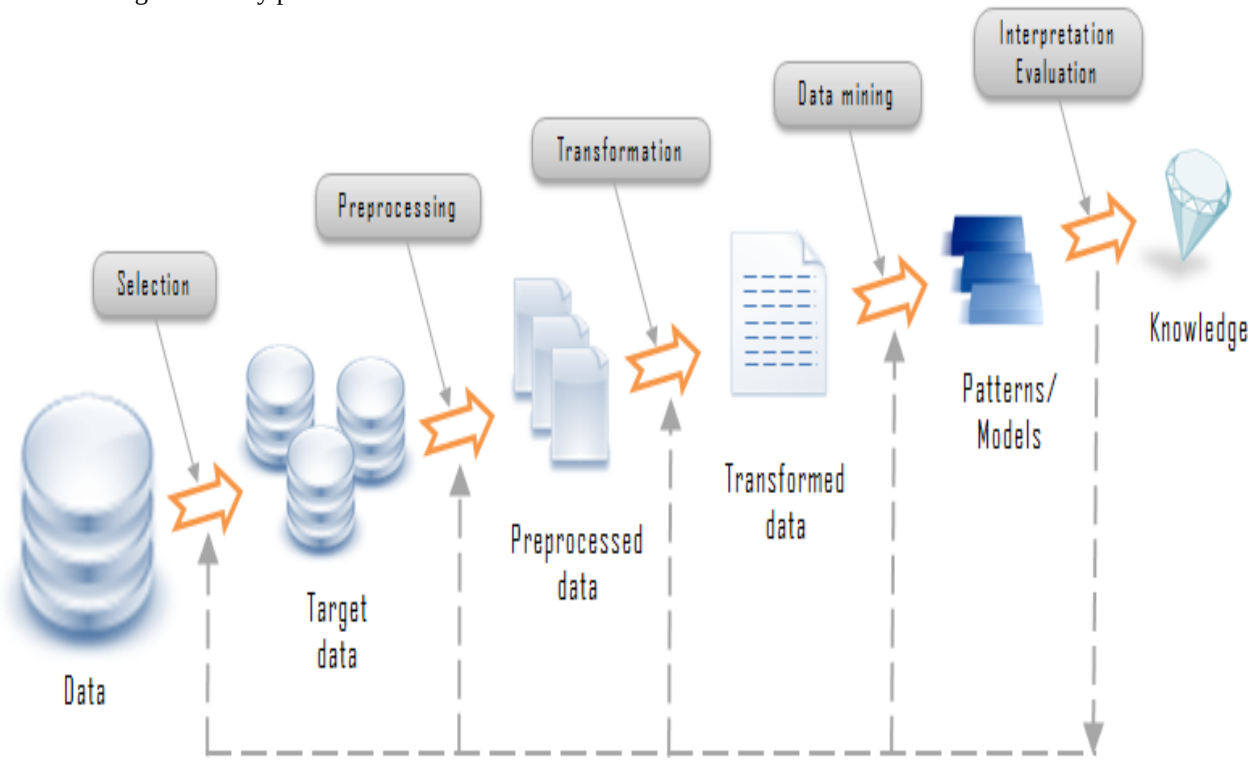
Keywords - Knowledge discovery is a process, Data mining Techniques

I. INTRODUCTION

The development of information technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.

II. KNOWLEDGE DISCOVERY PROCESS

Knowledge discovery is a process that extracts implicit, potentially useful or previously unknown information from the data. The knowledge discovery process is described as follows:



Let's examine the knowledge discovery process in the diagram above in details:

- Data comes from variety of sources is integrated into a single data store called target data

- Data then is pre-processed and transformed into standard format.
- The data mining algorithms process the data to the output in form of patterns or rules.
- Then those patterns and rules are interpreted to new or useful knowledge or information.

The ultimate goal of knowledge discovery and data mining process is to find the patterns that are hidden among the huge sets of data and interpret them to useful knowledge and information. As described in process diagram above, data mining is a central part of knowledge discovery process.

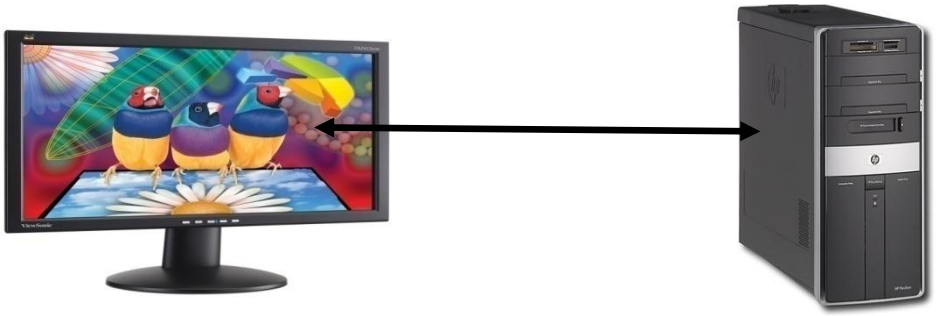
III. DATA MINING TECHNIQUES

There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns etc., are used for knowledge discovery from databases.

Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction.
For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.

Applications: market basket data analysis, cross-marketing, catalog design, loss-leader analysis, etc.



Types of association rules: Different types of association rules based on

- Types of values handled
  - → Boolean association rules
  - → Quantitative association rules
- Levels of abstraction involved
  - → Single-level association rules
  - → Multilevel association rules
- Dimensions of data involved
  - → Single-dimensional association rules
  - → Multidimensional association rules

Classification

Goal: Provide an overview of the classification problem and introduce some of the basic algorithms.
Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. For Example, Teachers classify students' grades as A, B, C, D, or F.

Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

In classification, we make the software that can learn how to classify the data items into groups. For example, we can apply classification in application that "given all past records of employees who left the company, predict which current employees are probably to leave in the future." In this case, we divide the employee's records into two groups that are "leave" and "stay". And then we can ask our data mining software to classify the employees into each group.

Classification Techniques

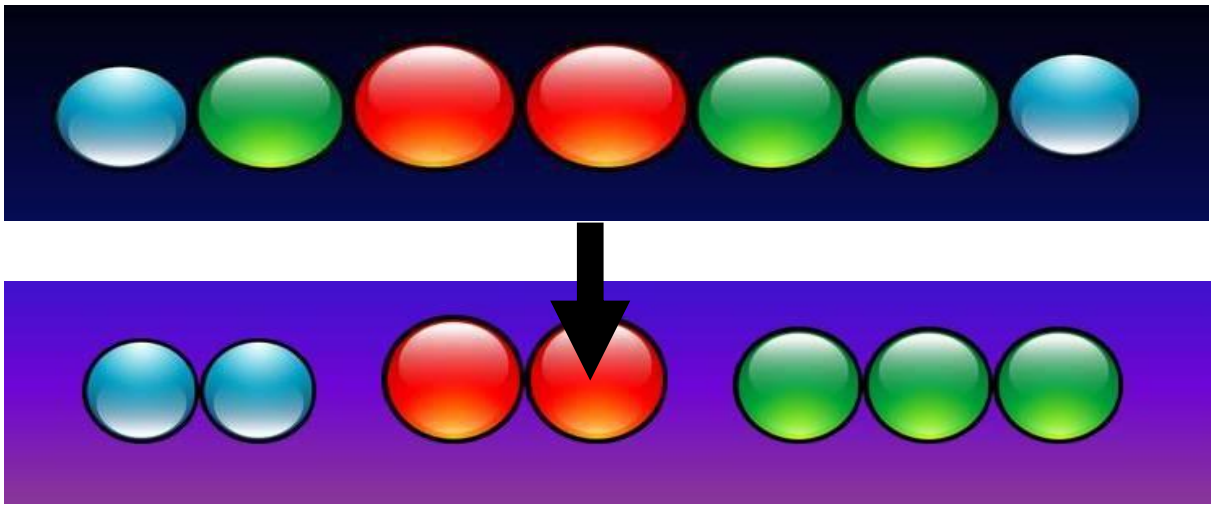
- Regression

- Distance
- Decision Trees
- Rules
- Neural Networks
-

- ***Clustering***

Clustering is "the process of organizing objects into groups whose members are similar in some way".

A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters



We can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without irritate. By using clustering technique, we can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in a topic, he or she would only go to that shelf instead of looking the whole in the whole library.

- ***Prediction***

The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables.

In data mining independent variables are attributes already known and response variables are what we want to predict unfortunately, many real-world problems are not simply prediction For instance, sales volumes, stock prices, and product failure rates are all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., decision trees) may be necessary to forecast future values.

For instance, prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

- ***Sequential Patterns***

Sequential patterns analysis in one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data.

- A sequence is an ordered list of events, denoted < e1e2… eL>.

- Each event $\mathbf{e_i}$ is an unordered set of items.
- Given two sequences α=< a1a2… an> and β=< b1b2… bm>
  α is called a subsequence of β, denoted as α⊆β, if there exist integers $1 \leq j1 < j2 < \ldots < jn \leq m$ such that a1⊆bj1, a2⊆bj2,…, an⊆bjn
  → Example: <a(bc)dc> is a *subsequence* of <a(abc)(ac)d(cf)>
- If a sequence contains *l*items, we call it a *l*-sequence
  → Example: <a(bc)dc> is a 5-sequence.
- The support of a sequence α is the number of data sequences that contain α.

IV. CONCLUSION

Data mining is a "decision support" process in which we search for patterns of information in data. In other words, Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc in different business domains. Data mining techniques such as classification, clustering, prediction, association and sequential patterns etc it helps in finding the patterns to decide upon the future trends in businesses to grow.

Data mining has wide application field almost in every industry where the data is generated that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology also.

REFERENCES

Journal Papers:

[1]. Dr. Lokanatha C. Reddy, A Review on Data mining from Past to the Future,
International Journal of Computer Applications (0975 – 8887) Volume 15– No.7, February 2011

[2]. Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth,From Data Mining to Knowledge Discovery in Databases, AI Magazine Volume 17 Number 3 (1996)

[3]. http://www.slideshare.net/Annie05/sequential-pattern-discovery-presentation

[4]. http://dataminingtools.net/wiki/introduction_to_data_mining.php

[5]. http://www.dataminingtechniques.net

[6]. http://www.slideshare.net/huongcokho/data-mining-concepts

Books:

[7]. Arun K. Pujari, *Data Mining Techniques*
[8]. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques

